

ADJUSTING STORM FREQUENCIES FOR CLIMATE FORECASTS

Thomas E. Croley II¹

ABSTRACT: Storm frequency estimates (e.g., maximum precipitation or flow probabilities) allow engineers and hydrologists to assess risks associated with their decisions during the design, construction, and operation of water resource projects. Storm frequencies for the future are often estimated directly from past historical records of sufficient length. The estimation requires no detailed knowledge of the area's meteorology, but presumes it remains unchanged in the future. However, the climate seldom remains static. Numerous climate forecasts of meteorology probabilities over extended periods are now available to the water resource engineer and hydrologist. It is possible to use these meteorology forecasts directly in the estimation of storm frequencies from the historical record. It is more desirable to do so now than at any time past, since meteorology forecasts have been improving and are now better than their predecessors. A heuristic approach is defined here to estimate storm frequencies that recognize forecasts of extended weather probabilities. Basically, those groups of historical meteorology record segments matching forecast meteorology probabilities are weighted more than others, during the estimation of storm frequencies. (Affiliated groups of hydrology record segments may be similarly weighted for hydrological estimation; e.g., flood frequency estimation.) An example frequency estimation of maximum flow is made using currently available agency meteorology forecasts in the US and Canada.

KEY TERMS: storms, exceedance frequencies, weather forecasts, probability, estimation, climate.

STORM FREQUENCY ESTIMATION

Since storm frequencies are unknown, they are *estimated* from the historical record, which is assumed ergodic and treated as a "random sample." Successive observations are considered identically distributed and equally likely to occur (both in the past and future). Likewise, the observations must be defined so they can be considered as independent of each other. (Two successive storms occurring very closely may result in a high degree of dependence of the second on the first.) Temporal dependence can be minimized by defining long event inter-arrival times or record pieces. For example, annual maximum floods or rainfalls (inter-arrival time on the order of a year) are often taken as time independent, as are 1-year record segments.

Storm frequencies or "exceedance probabilities," $P[X \geq x]$ can be estimated directly from the historical record. Suppose all values, x_i , in a random sample of annual maximums (x_i , $i = 1, \dots, n$) are ordered from largest to smallest to define the ordered variable values (y_ℓ , $\ell = 1, \dots, n$), where $y_\ell = x_{i(\ell)}$ and $i(\ell)$ is the number of the value in the unordered sample corresponding to the ℓ^{th} order. There are several methods to estimate exceedance probabilities from annual exceedance series (Chow, 1964); without loss of generality, the popular "Weibull" method is used here as an example:

$$\hat{P}[X \geq y_\ell] = \frac{\ell}{n+1} = \frac{1}{n+1} \sum_{i=1}^{\ell} 1, \quad \ell = 1, \dots, n \quad (1)$$

The caret, " $\hat{\cdot}$," denotes an estimate of the characteristic named underneath. Other methods also could be used.

This estimator is called "non-parametric" since knowledge of the underlying distribution and its parameters is not required. Other estimators (called "parametric") derive from knowledge (or supposition) of the type of underlying distribution. Functions of a random sample may be used as estimators of the parameters of the underlying distribution. Several of interest here are the "sample mean," $\hat{\mu}$, "sample variance," $\hat{\sigma}^2$, and "sample skew coefficient," $\hat{\psi}$:

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

¹Research Hydrologist, Great Lakes Environmental Research Laboratory, 2205 Commonwealth Blvd., Ann Arbor, Michigan 48105-2945, Phone: (734) 741-2238, Fax: (734) 741-2055, E-Mail: croley@glrl.noaa.gov.

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \hat{\mu})^2 \quad (3)$$

$$\hat{\psi} = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n (x_i - \hat{\mu})^3 / (\sqrt{\hat{\sigma}^2})^3 \quad (4)$$

They are estimators of distribution mean, μ , variance, σ^2 , and skew coefficient, ψ , respectively. Other estimators (Koutrouvelis and Canavos, 1999) also could be used with no loss of generality.

EXAMPLE STORM FREQUENCIES

The daily flow records for the Maumee River were searched for 1949-1995; the annual maximum daily flows are given in Table 1. The exceedance frequencies for the annual maximum flows were estimated with Eq. (1) and plotted in Figure 1 as "non-parametric sans forecast." The log-Pearson Type III distribution also was fit to the data and appears in Figure 1 as "parametric sans forecast." It results from supposing the natural logarithms of the data in Table 1 [$Z = \ln(X)$] are distributed as a three-parameter gamma distribution:

$$f_Z(z) = \frac{1}{|\beta| \Gamma(\alpha)} \left[\frac{z-c}{\beta} \right]^{\alpha-1} e^{-\frac{z-c}{\beta}}, \quad c \leq z < \infty \quad (\beta > 0) \quad -\infty < z \leq c \quad (\beta < 0) \quad (5)$$

where $f_Z(z)$ is probability density, $\Gamma(\alpha)$ is the gamma function, and α , β , and c are distribution parameters. Parameter estimates are given in terms of Eqs. (2)–(4) defined on the natural logarithms of the data (USWRC, 1967) by replacing expected values from Eq. (5) with sample moments:

$$\begin{aligned} \hat{\alpha} &= (2/\hat{\psi})^2 \\ \hat{\beta} &= \sqrt{\hat{\sigma}^2} \hat{\psi} / 2 \\ \hat{c} &= \hat{\mu} - 2\sqrt{\hat{\sigma}^2} / \hat{\psi} \end{aligned} \quad (6)$$

MATCHING PROBABILITY FORECASTS

The probability of any event A , $P[A]$, can be inferred with the estimator, $\hat{P}[A]$, defined as the number of observations in the random sample for which A occurs (i.e., for which the event A is true), n_A , divided by the total number of observations in the sample, n :

$$\hat{P}[A] = \frac{n_A}{n} = \frac{1}{n} \sum_{i|A} 1 \quad (7)$$

Table 1. Annual Maximum Daily Flow for the Maumee River Basin^a (35.31 cfs = 1 m³s⁻¹).

Year	Flow (cfs)	Year	Flow (cfs)	Year	Flow (cfs)	Year	Flow (cfs)
1949	45100	1961	53500	1973	40000	1985	91100
1950	92400	1962	45800	1974	69600	1986	36200
1951	53100	1963	35200	1975	49400	1987	23500
1952	53100	1964	46800	1976	68500	1988	22900
1953	33200	1965	36200	1977	64000	1989	42700
1954	23400	1966	79000	1978	86400	1990	82000
1955	45900	1967	48900	1979	53400	1991	86700
1956	42700	1968	56900	1980	44400	1992	54000
1957	62400	1969	67500	1981	85400	1993	65000
1958	29700	1970	33300	1982	113000	1994	63900
1959	80000	1971	38900	1983	54200	1995	51000
1960	44800	1972	46900	1984	51300		

^aat Waterville, Ohio, Lat. 41:30:00, Long. 83:42:46 (basin area = 16,394.7 km² = 6330 mi²).

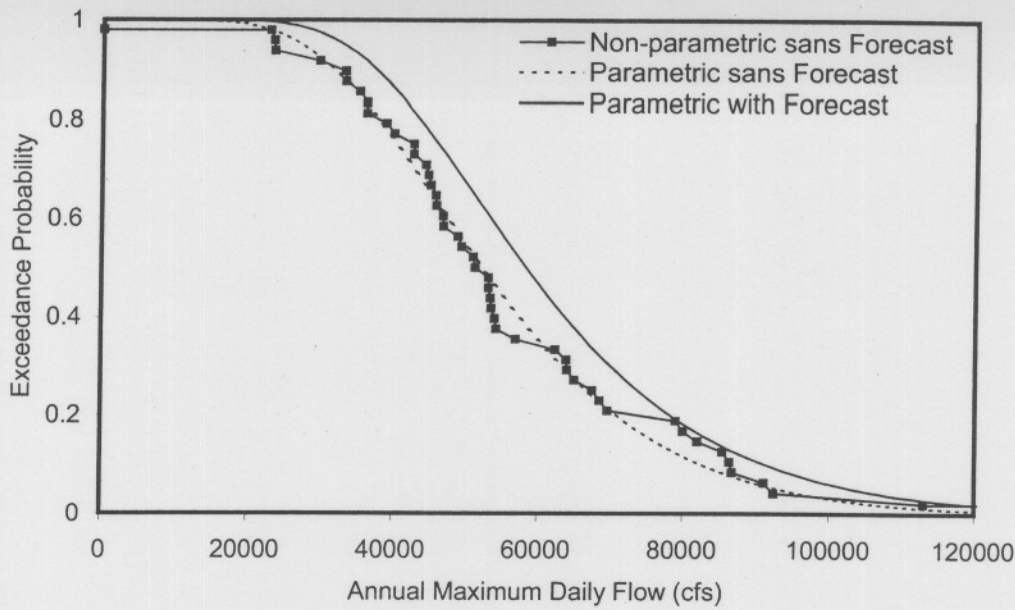


Figure 1. Annual Maximum Daily Maumee River Flow Exceedance Frequency (35.31 cfs = 1 m³s⁻¹).

In Eq. (7), the sum is taken over all i (members of the random sample) for which A occurs, denoted as $i|A$. The estimate in Eq. (7) is seen as the "relative frequency" of A in the random sample. Croley (1996, 1997, 2000) *biased* samples, by multiplying observations by non-negative weights, w_i , to calculate probabilities matching others' *multiple* probability forecasts:

$$\hat{P}[A] = \frac{1}{n} \sum_{i|A} w_i \quad (8)$$

$$\sum_{i=1}^n w_i = n \quad (9)$$

Consider, for example, that others' forecasts of event probability can be interpreted in $m-1$ probability equations (Croley, 1996) and forecasts of most-probable events can be interpreted in u probability inequalities (Croley, 2000). They are expressed in terms of relative frequencies over a random sample as follows:

$$\begin{aligned} \hat{P}[A_k] &= a_k, & k &= 2, \dots, m \\ \hat{P}[A_k] &\leq a_k, & k &= m+1, \dots, m+u \end{aligned} \quad (10)$$

where a_k are the forecast probabilities. Equation (8), when applied to match the forecasts of meteorology probabilities in Eq. (10) and added to Eq. (9), yield a system of equations to be solved for the weights:

$$\begin{aligned} \sum_{i=1}^n w_i &= n \\ \sum_{i|A_k} w_i &= n a_k, & k &= 2, \dots, m \\ \sum_{i|A_k} w_i &\leq n a_k, & k &= m+1, \dots, m+u \end{aligned} \quad (11)$$

Equivalently,

$$\begin{aligned} \sum_{i=1}^n \alpha_{k,i} w_i &= e_k, & k &= 1, \dots, m \\ \sum_{i=1}^n \alpha_{k,i} w_i &\leq e_k, & k &= m+1, \dots, m+u \end{aligned} \quad (12)$$

where $\alpha_{k,i} = 1$ for $k = 1$ or for $k > 1$ and $i \in A_k$ [inclusion of the i^{th} random sample value (i^{th} event or segment of the historical record) in the event of the k^{th} probability statement]; otherwise it is zero. Also, $e_k = n$ for $k = 1$ and $e_k = n a_k$ for $k > 1$. Any weights that satisfy Eq. (12) yield weighted-sample relative frequencies of events that match forecasts of meteorology probabilities. These weights also yield other corresponding biased sample estimators; e.g., Eqs. (1)–(4) become:

$$\begin{aligned}\hat{P}[X > y_\ell] &= \frac{1}{n+1} \sum_{k=1}^{\ell} w_{i(k)}, \quad \ell = 1, \dots, n \\ \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n w_i x_i \\ \hat{\sigma}^2 &= \frac{1}{n-1} \sum_{i=1}^n w_i (x_i - \hat{\mu})^2 \\ \hat{\psi} &= \frac{n}{(n-1)(n-2)} \sum_{i=1}^n w_i (x_i - \hat{\mu})^3 / (\sqrt{\hat{\sigma}^2})^3\end{aligned}\tag{13}$$

Generally, some of the equations in (11) or (12) may be either redundant or infeasible (non-intersecting with the rest, resulting in no solutions) and must be eliminated. (If the number of equations is greater than the number of weights, then some of the equations *must* be either redundant or infeasible.) In practice, one could assign each equation in (11) or (12) a “priority” reflecting its importance. [The highest priority is given to the first equation in (11) or (12) corresponding to Eq. (9), guaranteeing that all relative-frequencies sum to unity.] Each equation is compared to the set of all higher-priority equations and eliminated if redundant or infeasible. Thus Eq. (12) can always be reduced so that the allowed number of forecasts of meteorology probabilities is less than or equal to the number of historical record pieces (sample size). If less, then there are multiple solutions to Eq. (12), and a choice must be made as to which solution to use.

OPTIMUM SOLUTION

If there are multiple solutions to Eq. (12), the identification of the “best” requires a *measure* or *objective function* for comparing them. Solutions of Eq. (12) with larger values of this measure can be judged “better” than those with smaller values. One such measure is the probability of a selected event. If the objective function is always a statement of maximizing or minimizing a *probability*, then it can be added to the problem statement of Eq. (10) to yield an optimization problem. Objective functions that use probability statements can be expressed in the general form:

$$\max \sum_{i=1}^n \alpha_{0,i} w_i \tag{14}$$

where $\alpha_{0,i}$ are defined similarly to Eq. (12) in which the objective function is equation 0. The problem of solving Eq. (12) can now be formulated as an optimization, maximizing the objective function subject to a “constraint set” of equations:

$$\begin{aligned}\max \sum_{i=1}^n \alpha_{0,i} w_i \quad \text{subject to} \\ \sum_{i=1}^n \alpha_{k,i} w_i &= e_k, \quad k = 1, \dots, m \\ \sum_{i=1}^n \alpha_{k,i} w_i &\leq e_k, \quad k = m+1, \dots, m+u \\ w_i &\geq 0, \quad i = 1, \dots, n\end{aligned}\tag{15}$$

Equations (15) are amenable to standard “linear programming” optimization techniques. An algebraic procedure, termed the “Simplex” method, has been developed (Hillier and Lieberman, 1969) which progressively approaches the optimum solution through a well-defined iterative process until optimality is finally reached. Croley (2000) describes a procedure for applying the Simplex method in a two-stage optimization. The first stage finds a feasible solution to the constraint set in Eq. (15) and the second searches systematically from that feasible solution to the optimum solution. Multiple optima are possible, depending upon the objective function and the constraint set.

NOAA AND EC FORECAST EXAMPLE

The estimates of Figure 1 are modified by incorporating selected forecasts from the National Oceanic and Atmospheric Administration (NOAA event probability forecasts) and Environment Canada (EC most-probable event forecasts); see Croley (1996, 1997, 2000). The forecasts are summarized in Figure 2, in priority order with the earliest-made forecasts first (NOAA equalities precede EC inequalities), precipitation before temperature second, and chronologically third. In Figure 2, T_g = air temperature for period g and $\hat{t}_{g,\gamma}$ is the γ -quantile for period g air temperature estimated from a reference historical period (usually 1961-90 or 1963-93) such that:

$$\hat{P}[T_g \leq \hat{t}_{g,\gamma}] = \gamma \quad (16)$$

Likewise, precipitation, Q_g , and its quantiles, $\hat{\theta}_{g,\gamma}$, are defined similarly to Eq. (16).

Note that the precipitation forecasts in Figure 2 are for high precipitation with only one exception (the EC SON forecast). The objective in matching these forecasts is therefore (arbitrarily) taken as maximizing the probability that precipitation over the period November 1999–July 2000 will be in the upper third of its historical range (determined from 1961-1990):

$$\max \hat{P}[Q_{Nov'99-Jul'00} > \hat{\theta}_{Nov-Jul, 0.667}] \quad (17)$$

Daily precipitation and air temperature data from 32 stations were assembled over 1948-1995 for the Maumee River basin. (Widespread meteorology observations began in 1948.) The data were used to determine the Thiessen-averaged air temperature and the total precipitation, for the periods shown in Figure 2 and Eq. (17). According to the agencies, the NOAA temperature and precipitation forecasts and the EC precipitation forecasts are defined relative to historical reference quantiles estimated over the 1961-1990 period. Likewise, the EC temperature forecasts are defined relative to historical reference quantiles estimated over the 1963-1993 period. By ordering data from these periods, the reference quantiles are estimated.

Consider the objective function of Eq. (17) and the forecasts of Figure 2 to apply prior to and through the *beginning* of each year in the sample. (The Maumee River annual maximum flow typically occurs as spring snowmelt.) In other words, each year of record is to be weighted to reflect the objective of Eq. (17) and the beginning winter as forecast in Figure 2 (a total period from September of the year before through the following August). For example, the first value in Table 1 for calendar year 1949 corresponds to the objective and forecast values for September 1948–August 1949. The coefficients in Eq. (15) are derived from the data set, Figure 2, and Eq. (17); see Croley (2000). In the ensuing optimization, 19 weights are zeroes, indicating that some of the historical record is not used. However, all but the last three equations in Figure 2 are used (corresponding to all forecasts except the EC most-probable JJA air temperature forecast).

Climate-biased storm frequencies for the annual maximum daily flow can now be estimated by applying these weights to the data in Table 1 by using Eq. (13). Only results for the fitted Log-Pearson Type III distribution are given in Figure 1 (to simplify the presentation) as “parametric with forecasts.” Compare the Log-Pearson Type III distribution derived from the parametric estimates without the forecasts to that made with the forecasts. There is a large shift, making all flows more likely to be exceeded.

SUMMARY AND OBSERVATIONS

The methodology described herein allows one to recognize changing climate in the estimation of storm frequencies, removing one of the worst assumptions associated with this, which is that future probabilities are the same as the past. Existing forecasts of meteorology probabilities can be used to bias storm frequency estimates for a changing climate. The methodology is adapted from earlier work that uses forecasts of meteorology probabilities to derive forecasts of consequent hydrology probabilities in an operational hydrology approach. The linear objective function used here enables incorporation of an event probability into the objective, use of existing optimization techniques, and direct inclusion of non-negativity constraints.

The example presented here may be more representative of storm frequency estimation in an *operational* setting rather than in a *design* setting. Climate-biased storm frequencies were estimated by *preserving* meteorology forecasts. These conditions are current and are not generally regarded as applying over a very long time into the future. The resulting biased storm frequencies can only be considered applicable over the same time period as the meteorology forecasts or other event probabilities used to condition them. The example given here applied over the next several months, appropriate for use in an operational setting. If probabilities can be defined (estimated) corresponding to climate shifts expected from the present forward, then the resulting biased storm frequencies could be used in a design setting.

Complete software, in the form of an easy-to-use interactive *Windows*™ graphical user interface, and worked examples are available free of charge over the World Wide Web. The software, examples, and tutorial materials may be acquired in a self-installing file by visiting the web site entitled: <http://www.glerl.noaa.gov/wr/OutlookWeights.html> and downloading.

$\hat{P}[Q_{OND}'99 \leq \hat{\theta}_{OND, 0.333}] = 0.283$	(1)	$\hat{P}[T_{FMA}'00 \leq \hat{t}_{FMA, 0.333}] = 0.333$	(21)
$\hat{P}[Q_{OND}'99 > \hat{\theta}_{OND, 0.667}] = 0.383$	(2)	$\hat{P}[T_{FMA}'00 > \hat{t}_{FMA, 0.667}] = 0.333$	(22)
$\hat{P}[Q_{NDJ}'99 \leq \hat{\theta}_{NDJ, 0.333}] = 0.273$	(3)	$\hat{P}[T_{MAM}'00 \leq \hat{t}_{MAM, 0.333}] = 0.333$	(23)
$\hat{P}[Q_{NDJ}'99 > \hat{\theta}_{NDJ, 0.667}] = 0.393$	(4)	$\hat{P}[T_{MAM}'00 > \hat{t}_{MAM, 0.667}] = 0.333$	(24)
$\hat{P}[Q_{DJF}'99 \leq \hat{\theta}_{DJF, 0.333}] = 0.273$	(5)	$\hat{P}[Q_{SON}'99 \leq \hat{\theta}_{SON, 0.333}] \leq 0.333$	(25)
$\hat{P}[Q_{DJF}'99 > \hat{\theta}_{DJF, 0.667}] = 0.393$	(6)	$\hat{P}[\hat{\theta}_{SON, 0.333} < Q_{SON}'99 \leq \hat{\theta}_{SON, 0.667}] \geq 0.334$	(26)
$\hat{P}[Q_{JFM}'00 \leq \hat{\theta}_{JFM, 0.333}] = 0.133$	(7)	$\hat{P}[Q_{SON}'99 > \hat{\theta}_{SON, 0.667}] \leq 0.333$	(27)
$\hat{P}[Q_{JFM}'00 > \hat{\theta}_{JFM, 0.667}] = 0.533$	(8)	$\hat{P}[Q_{JJA}'00 \leq \hat{\theta}_{JJA, 0.333}] \leq 0.333$	(28)
$\hat{P}[Q_{FMA}'00 \leq \hat{\theta}_{FMA, 0.333}] = 0.273$	(9)	$\hat{P}[\hat{\theta}_{JJA, 0.333} < Q_{JJA}'00 \leq \hat{\theta}_{JJA, 0.667}] \leq 0.334$	(29)
$\hat{P}[Q_{FMA}'00 > \hat{\theta}_{FMA, 0.667}] = 0.393$	(10)	$\hat{P}[Q_{JJA}'00 > \hat{\theta}_{JJA, 0.667}] \geq 0.333$	(30)
$\hat{P}[Q_{MAM}'00 \leq \hat{\theta}_{MAM, 0.333}] = 0.273$	(11)	$\hat{P}[T_{SON}'99 \leq \hat{t}_{SON, 0.333}] \leq 0.333$	(31)
$\hat{P}[Q_{MAM}'00 > \hat{\theta}_{MAM, 0.667}] = 0.393$	(12)	$\hat{P}[\hat{t}_{SON, 0.333} < T_{SON}'99 \leq \hat{t}_{SON, 0.667}] \leq 0.334$	(32)
$\hat{P}[T_{OND}'99 \leq \hat{t}_{OND, 0.333}] = 0.333$	(13)	$\hat{P}[T_{SON}'99 > \hat{t}_{SON, 0.667}] \geq 0.333$	(33)
$\hat{P}[T_{OND}'99 > \hat{t}_{OND, 0.667}] = 0.333$	(14)	$\hat{P}[T_{DJF}'99 \leq \hat{t}_{DJF, 0.333}] \leq 0.333$	(34)
$\hat{P}[T_{NDJ}'99 \leq \hat{t}_{NDJ, 0.333}] = 0.333$	(15)	$\hat{P}[\hat{t}_{DJF, 0.333} < T_{DJF}'99 \leq \hat{t}_{DJF, 0.667}] \leq 0.334$	(35)
$\hat{P}[T_{NDJ}'99 > \hat{t}_{NDJ, 0.667}] = 0.333$	(16)	$\hat{P}[T_{DJF}'99 > \hat{t}_{DJF, 0.667}] \geq 0.333$	(36)
$\hat{P}[T_{DJF}'99 \leq \hat{t}_{DJF, 0.333}] = 0.273$	(17)	$\hat{P}[T_{JJA}'00 \leq \hat{t}_{JJA, 0.333}] \geq 0.333$	(37)
$\hat{P}[T_{DJF}'99 > \hat{t}_{DJF, 0.667}] = 0.393$	(18)	$\hat{P}[\hat{t}_{JJA, 0.333} < T_{JJA}'00 \leq \hat{t}_{JJA, 0.667}] \leq 0.334$	(38)
$\hat{P}[T_{JFM}'00 \leq \hat{t}_{JFM, 0.333}] = 0.263$	(19)	$\hat{P}[T_{JJA}'00 > \hat{t}_{JJA, 0.667}] \leq 0.333$	(39)
$\hat{P}[T_{JFM}'00 > \hat{t}_{JFM, 0.667}] = 0.403$	(20)		

Figure 2. Mixed NOAA and EC Meteorology Probability Forecasts Made in September 1999 over the Maumee River Basin.

ACKNOWLEDGMENTS

This is GLERL contribution no. 1151.

REFERENCES

- Chow, V. T., 1964. *Handbook of Applied Hydrology*. McGraw-Hill, New York, 8-29.
- Croley, T.E., II, 1996. "Using NOAA's New Climate Outlooks In Operational Hydrology." *Journal of Hydrologic Engineering*, ASCE, 1(3):93-102.
- Croley, T.E., II, 1997. Mixing Probabilistic Meteorology Outlooks In Operational Hydrology, *Journal of Hydrologic Engineering*, ASCE, 2(4):161-168.
- Croley, T. E., II, 2000. *Using Meteorology Probability Forecasts In Operational Hydrology*. ASCE Press, American Society of Civil Engineers, Reston, Virginia (in press).
- Hillier, F. S., and Lieberman, G. J., 1969. *Introduction to Operations Research*. Chapter 5: Linear Programming, Holden-Day, San Francisco, 127-171.
- Koutrouvelis, I. A., and Canavos, G. C., 1999. "Estimation In The Pearson Type 3 Distribution." *Water Resource Research*, AGU, 35(9):2693-2704.
- U.S. Water Resources Council, 1967. A Uniform Technique For Determining Flood-Flow Frequency. *Bulletin 15*, Hydrology Committee, Washington, DC, 1967.